

Respondent-Driven Sampling in a Study of Drug Users in New York City: Notes from the Field

Courtney McKnight, Don Des Jarlais, Heidi Bramson, Lisa Tower,
Abu S. Abdul-Quader, Chris Nemeth, and Douglas Heckathorn

ABSTRACT *Beth Israel Medical Center (BIMC), in collaboration with the Centers for Disease Control (CDC) and the New York State Department of Health (NYSDOH), used respondent-driven sampling (RDS) in a study of HIV seroprevalence among drug users in New York City in 2004. We report here on operational issues with RDS including recruitment, coupon distribution, storefront operations, police and community relations, and the overall lessons we learned. Project staff recruited eight seeds from a syringe exchange in Lower Manhattan to serve as the initial study participants. Upon completion of the interview that lasted approximately 1 h and a blood draw, each seed was given three coupons to recruit three drug users into the study. Each of the subsequent eligible participants was also given three coupons to recruit three of their drug-using acquaintances. Eligible participants had to have: injected, smoked or snorted an illicit drug in the last 6 months (other than marijuana), aged 18 or older, adequate English language knowledge to permit informed consent and complete questionnaire. From April to July 2004, 618 drug users were interviewed, including 263 (43%) current injectors, 119 (19%) former injectors, and 236 (38%) never injectors. Four hundred sixty nine (76%) participants were men, 147 (24%) were women, and two (<1%) were transgender. By race/ethnicity, 285 (46%) were black, 218 (35%) Hispanic, 88 (14%) white, 23 (4%) mixed/not specified, and four (<1%) native American. Interviews were initially done on a drop-in basis but this system changed to appointments 1 month into the study due to the large volume of subjects coming in for interviews. Data collection was originally proposed to last for 1 year with a target recruitment of 500 drug users. Utilizing RDS, we were able to recruit and interview 118 more drug users than originally proposed in one quarter of the time. RDS was efficient with respect to time and economics (we did not have to hire an outreach worker) and effective in recruiting a diverse sample of drug users.*

KEYWORDS CAPI, Drug users, IRIS Plus, Respondent-driven sampling (RDS), Time-space sampling.

BACKGROUND

A variety of sampling methods have been used to recruit hard-to-reach populations, such as drug users, into research studies. These methods include snowball sampling,

McKnight, Des Jarlais, Bramson, and Tower are with The Baron Edmond de Rothschild Chemical Dependency Institute, Beth Israel Medical Center, New York, NY; Abdul-Quader is with the Centers for Disease Control and Prevention, Atlanta, GA, USA; Nemeth is with the New York State Department of Health, Albany, NY, USA; Heckathorn is with the Cornell University, Ithaca, NY, USA.

Correspondence: Courtney McKnight, The Baron Edmond de Rothschild Chemical Dependency Institute, Beth Israel Medical Center, New York, NY. (E-mail: cmcknight@chpnet.org)

targeted sampling, and time-space sampling.¹⁻³ Despite widespread use of these methods, they are not without important limitations.

Snowball sampling begins with a set of initial study participants, called seeds, who refer other eligible respondents. Once these persons are interviewed, they are also asked to bring in or provide references for other potential respondents. This process continues until the sample size is achieved. Snowball sampling can provide easy access to hidden populations, but it is most often biased because it is usually done out of convenience rather than randomly.¹⁻³

Targeted sampling involves thoroughly collecting preliminary data to determine various characteristics such as locations where a sample congregates and demographic characteristics such as race, gender, and age.^{1,2,4} Similar to snowball sampling, using targeted sampling to study drug users may result in a selection bias because less-visible, isolated drug users may not be found at the sites where other drug users are sampled and, thus, may be overlooked.

Time-space sampling involves preliminary research by ethnographers to determine when and where hidden populations congregate.^{3,5} Once a list of times and sites is determined, researchers then randomly choose times to visit those sites to collect data. Time-space sampling has limited use with drug users because making an exhaustive list of where drug users congregate could be a very protracted and expensive process, as drug-use sites change frequently in response to environmental factors such as police presence.

Respondent-driven sampling (RDS),¹ a relatively new method being used to draw probability samples of hidden populations, incorporates some methods of snowball sampling, such as chain referral sampling, but includes additional provisions to minimize bias. It produces population estimates that are asymptotically unbiased, which means that bias is only on the order of $1/[\text{sample size}]$, so bias is minimal in samples of substantial size.⁶ This allows for probability-based inferences based on the social network of the sample.

This paper reports on the use of RDS to recruit participants for a study of risk behavior and HIV seroprevalence among drug users in New York City. The study was conducted by Beth Israel Medical Center in collaboration with the Centers for Disease Control and Prevention and the New York State Department of Health. We discuss the operational issues faced with RDS such as recruitment, coupon distribution, storefront operations, police and community relations, and the overall lessons learned.

METHODS

In April 2004, Beth Israel Medical Center staff recruited eight seeds from a syringe exchange in Lower Manhattan. Each seed was screened for eligibility before being given a coupon to come back for an interview. The seeds were asked about their drug use, including mode of use and drug preparation techniques. Those who claimed to be an injection drug user (IDU) were also asked to show track marks. Although not a requirement of RDS, the seeds were recruited to resemble the race and gender profile of drug users in Lower Manhattan.⁷ They were asked to come to a research storefront in Lower Manhattan the following day to complete a computer-assisted interviewer-administered personal interview (CAPI) and to have their blood drawn for an HIV test. Each seed, and subsequent study subject, received \$20 compensation for their time.

When subjects arrived at the storefront, they were questioned by the study screener to ensure eligibility. Once they were deemed eligible by the screener, each subject was assigned a unique code to serve as their study identification. This code included the following information: the first two letters of the last name, first letter of mother's first name, the last two digits of the birth year, one letter for the person's race, and one letter for gender. This code was used to identify blood work and questionnaires and could be regenerated if the subject forgot it. No names or other identifying information were asked.

After being assigned a study code, subjects met with an interviewer to be consented and interviewed. The interview had three parts, an hour-long questionnaire, administered by the interviewer, HIV pre-test counseling and a blood draw for an HIV test. The interview consisted of a structured questionnaire which took approximately 1 h and asked about drug-use frequency, drug and sexual risk behavior, syringe acquisition, and knowledge of HIV and hepatitis B and C. After the interview, HIV counseling, and blood collection, each subject was given three coupons to recruit three other drug users into the study. The subjects were briefly trained on how to recruit others, with specific emphasis on the recruitment of friends and acquaintances who use drugs. Eligible respondents had to have done the following: injected, smoked, or snorted an illicit drug in the past 6 months (those who smoked only marijuana were not eligible for the study); turned age 18 years or older by the time of the interview; been able to speak English adequately to consent to the study and complete the questionnaire; and lived, bought, and/or used drugs on the Lower East Side of Manhattan. Subjects could have, however, lived in the larger New York metropolitan area, as long as they bought or used their drugs on the Lower East Side of Manhattan.

Coupons contained the time that the storefront opened (9:00 A.M.) and a unique number (to make each one distinct). To make them difficult to duplicate, coupons were printed on thick cardstock paper with color images. To track the coupons and payment for each respondent, we used custom-developed software for RDS called IRIS Plus. Information such as respondent's unique code, physical traits, coupon number, and the numbers of the coupons each respondent distributed were all recorded in IRIS Plus. This information enabled us to link coupons together, determine when respondents should be paid and who gave coupons to whom. This software also helped to prevent the redemption of duplicated coupons because the database would not accept duplicate coupon numbers. When respondents came in to make an appointment, the coupon was checked in the IRIS Plus database to verify that it had not been used previously. Additionally, each person's code was checked in IRIS Plus after being screened to determine whether a person with that code had previously been enrolled in the study. If another study subject had that code, the screener looked at the physical traits listed for that person for verification. Additionally, if the screener felt that a person looked familiar, she would search for the person's physical traits in IRIS Plus to see if another person with similar characteristics was previously enrolled.*

On July 2, 2004, we ceased coupon distribution to give potential subjects 2 weeks to redeem remaining coupons. On July 16, 2004, we concluded data collection, and the study officially concluded on July 30, 2004. The last 2 weeks of

*We had one instance of a person trying to re-enroll in the study. Even though the person tried to disguise himself with a hat and glasses, the screener recognized him and did not enroll him in the study.

July were set aside to allow the remainder of subjects to come in for their HIV test results.

RESULTS

From April to July 2004, 618 respondents were interviewed, including 263 (43%) current injectors (had injected drugs within the past 6 months), 119 (19%) former injectors (had injected drugs more than 6 months prior to the interview—who used drugs in the past 6 months by other modes of administration), and 236 (38%) never injectors (those that had never injected any drugs but used non-injection drugs in the past 6 months). The mean age of respondents was 44 years, with a mean age of first drug use at 19 years and mean age of first injection drug use at 22 years (among current and former drug injectors). By gender, 469 (76%) were men, 147 (24%) were women, and two (<1%) were transgender. By race/ethnicity, 285 (46%) were black; 218 (35%), Hispanic; 88 (14%), white; 23 (4%), mixed race or unspecified race; and four (<1%), native American.

Seed Recruitment

To control the flow of recruitment, we recruited only eight seeds: two white men, one white woman, two Hispanic men, one Hispanic woman, one black man, and one black woman, all of whom identified themselves as active IDUs. A relatively small number of seeds was recruited so that we could assess how productive they were and then recruit more if necessary, without being too overwhelmed with respondents.

Respondent Recruitment

Our initial goal was to interview 500 drug users within 1 year. After 4 months, 618 drug users had been interviewed. Respondent recruitment (those people recruited after the seeds) happened much faster than we had expected. In the first couple of weeks, recruitment was slow but quickly picked up as more coupons got out into the community. Interviews were initially done on a drop-in basis, but this system changed to appointments after 1 month into the study due to the large volume of respondents coming in for interviews. As the flow of potential subjects continued to be rapid and at times overwhelming, we were also forced to post-date all coupons by 2 days. Therefore, when a respondent was given a coupon, that person could not be interviewed or make an appointment until the date written on the coupon. This system was helpful in pacing the number of people coming into the storefront for interview appointments but perhaps choosing a smaller number of seeds may have alleviated the need to implement these changes altogether.

Police and Community Relations

As previously noted, as the study progressed and more coupons were distributed in the community, the number of people coming into the storefront for interviews increased dramatically. For the first 2 weeks, the interviewers would arrive in the morning to find 10 to 20 people waiting for them. People reported arriving as early as 7:00 A.M. so that they could be interviewed that day. Once crowds began to form outside the storefront, neighboring businesses complained to our staff or called the police to file complaints. The police came to speak with staff on three separate occasions regarding complaints of noise and loitering. At that point we decided to move to an all-appointment system. Instead of telling people that their recruits

could show up and be interviewed, recruits were told to first come into the storefront to be screened, and that if they passed the screener, they could make an appointment.

Lessons Learned

Were we to do another RDS study in New York City, we would do the following:

We would have a phone number so that screening and appointments for potential participants could be done by telephone. This would lessen the amount of traffic in the storefront and reduce some confusion. Once participants arrived for their interviews, we would then screen them about their drug use.

If we were looking to recruit a large number of drug users in a short amount of time, we would again recruit eight seeds. The eight seeds recruited for this study seemed to be an appropriate number because we were trying to maximize our time in the field for data collection. Although the first 2 weeks of the study were relatively slow-paced, it gave the field staff time to adjust to their new environment and new roles and make any necessary changes before they became too overwhelmed. Drug users in New York City are generally a very overstudied population, and thus we expected a large portion of the seeds to show up for their interviews. In cities where the population is not as familiar with research, researchers may want to recruit more seeds, considering that many may not return for the initial interview. Another alternative, however, if time permits, is to recruit a smaller number of seeds, give them a few weeks to return for their interview and then recruit more seeds, if necessary. It was our intention to recruit more seeds if our initial eight were not productive, but this was not necessary.

We would start with an appointment-only system. We began with a drop-in system for interviews because our time to collect data was very limited and this was the most efficient use of the interviewers' time. For instance, if a person dropped in and wanted to be interviewed, we could do it on the spot if the person with a scheduled interview had not shown up. Even though this system worked for us initially, it soon became unwieldy, and we were forced to change a month into the study. People adjusted very quickly, and many people reported being relieved because they didn't have to wait in front of the storefront for 2 h in the morning before staff arrived.

We would be very rigid about our appointments. In this study we were very strict about appointment times and conveyed this to each participant. If persons with appointments were more than 5 min late and someone else arrived to schedule an appointment but was free to complete it then, we took the person who dropped in rather than waiting for the person with the scheduled appointment to show up. We rarely had interviewers sitting around with nothing to do. There was almost always someone who dropped in when someone else was late for an appointment.

We would meet with local business owners to inform them of who we are and what we were doing prior to the start of the study. We would invite them to tell us if they had any concerns. We would not disclose the type of research being done or with whom, but it would be helpful to let the surrounding business owners know that they may see an increase in the number of people in or around the storefront.

We would remain at the storefront 6 weeks after ending coupon distribution. Coupon distribution ended a full 4 weeks before we vacated the storefront, and data collection ended 2 weeks before. The timing of both of these allowed for a fair amount of those with coupons to redeem them, for those with outstanding incentives to collect their money, and for those with outstanding test results to come

in for them. Two additional weeks would have been helpful to allow for remaining coupons to be redeemed. Because the storefront we used stayed open for other research studies, we were able to leave contact information for our interviewers in case any respondents showed up for their HIV test results. Within the month after we left the storefront, we received two additional calls for HIV test results. Our interviewers met the respondents at a park and a coffee shop to give them their results. The other storefront staff reported a handful of people coming in with coupons after the study closed, but they did not report any incidents with those people that were unable to redeem them. Finally, we did not pay out any uncollected incentives after July 30, 2004.

Some of the lessons that we learned about conducting a study using RDS may be specific to New York City. Because there was an abundance of willing and able participants, we had to take several measures to make data collection more manageable. These measures may not be necessary or practical for researchers in cities where the drug-using population is smaller, less willing to participate in research, or both.

CONCLUSIONS

Using RDS, we were able to recruit and interview 118 more drug users than originally proposed in one third of the time. In our experience, RDS was efficient with respect to time and economics (we did not have to hire an outreach worker) and effective in garnering a diverse sample of drug users. We were able to interview many more drug users than expected in a short period of time with very minimal recruitment effort. In summary, RDS can be an efficient and effective form of recruitment, particularly for research involving drug users or other hidden populations.

REFERENCES

1. Heckathorn D. Respondent-driven sampling: a new approach to the study of hidden populations. *Soc Probl.* 1997;44:174–199.
2. Heckathorn D. Respondent driven sampling II. Deriving valid population estimates from chain-referral samples of hidden populations. *Soc Probl.* 2002;49:11–34.
3. Magnani R, Sabin K, Saidel T, Heckathorn D. Review of sampling hard-to-reach and hidden populations or HIV surveillance. *AIDS.* 2005;19(Suppl 2):S67–S72.
4. Watters JK, Biernacki P. Targeted sampling: options for the study of hidden populations. *Soc Probl.* 1989;36:416–430.
5. Semaan S, Lauby J, Liebman J. Street and network sampling in evaluation studies of HIV risk-reduction interventions. *AIDS Rev.* 2002;4:213–223.
6. Salganik MJ, Heckathorn D. Sampling and estimation in hidden populations using respondent-driven sampling. *Sociol Method.* 2004;34:193–239.
7. Des Jarlais DC, Diaz T, Perlis T, et al. Variability in the incidence of human immunodeficiency virus, hepatitis B virus, and hepatitis C virus infection among young injecting drug users in New York City. *Am J Epidemiol.* 2003;157(5):467–471.